

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Centrality as a gradual notion: A new bridge between fuzzy sets and statistics

Pedro Terán

Escuela Politécnica de Ingeniería, Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, E-33071 Gijón, Spain

ARTICLE INFO

Article history:

Available online 27 May 2011

Keywords:

Centrality
Choquet integral
Fuzzy event
Multivariate analysis
Random set

ABSTRACT

This paper aims at formalizing the intuitive idea that some points are more central in a probability distribution than others. Our proposal relies on fuzzy events to define a fuzzy set of central points for a distribution (or a family of distributions, including imprecise probability models). This framework has a natural interpretation in terms of fuzzy logic and unifies many known notions from statistics, including the mean, median and mode, interquantile intervals, the Lorenz curve, the halfspace median, the zonoid and lift zonoid, the coverage function and several expectations and medians of random sets, and the Choquet integral against an infinitely alternating or infinitely monotone capacity.

© 2011 Published by Elsevier Inc.

1. Introduction

As evidenced by [10, Sections 1 and 2], Fuzzy Set Theory and Statistics have a consistent record of mutual misunderstandings. Since that fire remains unextinguished, as recent papers (see [28,32]) and the continued citation flow of controversial old work (e.g. [21]) seem to demonstrate, developing bridges between both cultures looks no less important today than it ever was. Some recent work in that direction can be found in [27,6,8,26,11,39,7,12,9].

This paper tries to contribute by exploring the consequences of regarding centrality in a distribution, the key notion in location estimation, as a gradual notion. Paraphrasing a famous sentence of Orwell's, our motto might be 'All points are central, but some points are more central than others'. The basic object in this framework is the fuzzy set of central points of the distribution or distributions at hand. The membership degrees in the fuzzy set of central points admit an interesting fuzzy logic interpretation.

Most of the content of the paper goes toward showing that many familiar notions from statistics can be retrieved within that framework by looking at the maximally central points. We consider both the case of a single probability and models involving a whole family of distributions. The underlying space is \mathbb{R}^d , thus covering multivariate as well as univariate data.

The key underlying concepts are fuzzy events and their probabilities. A *fuzzy event* is a measurable mapping from a measurable space to $[0, 1]$. Each ordinary (Boolean) event B is identified with its indicator function I_B . Following Zadeh [37], a probability measure P on Boolean events extends to fuzzy events by the formula

$$P(A) = \int A dP.$$

The paper is structured as follows. Section 2 presents the basic definitions and examples. Section 3 justifies the convenience of extending the framework to families of distributions instead of single distributions. Most findings in these are summarized in the abstract. Sections 4 and 5 work out the specific application to random sets and Choquet capacities. The former shows that the Aumann, Herer and Vorob'ev expectations, the Vorob'ev median and the coverage function of a random set

E-mail address: teranpedro@uniovi.es

are subsumed by our framework. As a consequence, the zonoid and lift zonoid (which serves as a multivariate version of the Lorenz curve) are included too. In the latter section, the fuzzy set of central points of a Choquet capacity is given a fuzzy logic interpretation, and the Choquet integral is shown to arise from another special case. Finally, some interesting aspects are discussed in Section 6.

2. A fuzzy notion of centrality

Let P be the probability distribution of a random variable or vector. Consider the following questions:

1. Which point(s) are most central with respect to P ?
2. How to measure the centrality in P of a given point?

By giving each point a degree of centrality, one defines a *fuzzy set of central points*. Then, a point with the maximal degree of centrality would be a natural location estimator. The first problem is how to elicit the degree of centrality of a point in a way consistent with ordinary statistical practice, so that common location estimators become maximally central points.

In this section, we introduce a general framework to deal with this problem. It is intuitively appealing and allows one to give a unified definition of the expectation, the median and the mode as examples of maximally central points. In further sections, by extending the notion of centrality to families of distributions (instead of the single distribution of a random variable), we will show how to move beyond the location problem and how to encompass imprecise probability models.

2.1. Definitions

The core intuition is as follows. If, of two points x and y , the former lies closer to the center of a distribution, while the latter is outlying, we would wish to be able to say, somehow, that events around x tend to have larger probabilities than those containing y . Still, that intuition seems to be flawed in that (a) it relies on our familiarity with unimodal distributions, (b) it is false if taken literally: for any event A containing x , $A \cup \{y\}$ differs from A by at most one point, contains y , and has the same or a larger probability.

Is it possible to formalize that intuition while avoiding those pitfalls? Since allowing arbitrary events (like those generated by adding one point to another event) will not do, we must restrict the events under consideration. Thus the first ingredient is a family \mathcal{A} of privileged *reference events*. That family will be formed by fuzzy events of the space \mathbb{R}^d where the random variable ($d = 1$) or vector ($d > 1$) takes on values. Different reference families may lead to different notions of centrality and associated location estimators.

Roughly speaking, a point x will be called central if all reference events containing x will probably happen. That is reminiscent of the mode as the most likely point of a discrete distribution. In a more formal fashion, we call x *central* in P (with respect to the family \mathcal{A}) if

$$\forall A \in \mathcal{A}, \quad P(A) \geq A(x).$$

But although this may capture the notion that the value of $P(A)$ is required to be large insofar as the degree to which x belongs to A is large, a gradual restriction seems even more faithful. Thus, x is called α -central (with respect to \mathcal{A}) if

$$\forall A \in \mathcal{A}, \quad P(A) \geq \alpha \cdot A(x).$$

Since the sets of α -central points are nested, that defines a *fuzzy set of central points* by the formula

$$C(x) = \sup\{\alpha \mid x \text{ is } \alpha\text{-central in } P\}.$$

For clarity of exposition, the notation omits the dependence of C on the choice of \mathcal{A} . The height of C will be denoted $h(C)$.

That definition can be rewritten in the language of fuzzy logic as follows. Let I denote Goguen's fuzzy implication, namely $I: [0, 1] \times [0, 1] \rightarrow [0, 1]$ is the function given by

$$I(x, y) = \begin{cases} y/x, & x \geq y \text{ and } x \neq 0, \\ 1, & \text{otherwise.} \end{cases}$$

Proposition 2.1. Let P be a probability distribution in \mathbb{R}^d , \mathcal{A} a family of reference events and $x \in \mathbb{R}^d$. Then,

$$C(x) = \inf_{A \in \mathcal{A}} I(A(x), P(A)).$$

Proof. We have

$$\begin{aligned} C(x) &= \sup \{ \alpha \in (0, 1] \mid P(A) \geq \alpha \cdot A(x) \forall A \in \mathcal{A} \mid A(x) > 0 \} = \sup \left\{ \alpha \in (0, 1] \mid \frac{P(A)}{A(x)} \geq \alpha \forall A \in \mathcal{A} \mid A(x) > 0 \right\} \\ &= \sup \left\{ \alpha \in (0, 1] \mid \inf_{A \in \mathcal{A} \mid A(x) > 0} \frac{P(A)}{A(x)} \geq \alpha \right\} = \min \left\{ \inf_{A \in \mathcal{A} \mid A(x) > 0} \frac{P(A)}{A(x)}, 1 \right\} = \inf_{A \in \mathcal{A} \mid A(x) > 0} \min \left\{ \frac{P(A)}{A(x)}, 1 \right\} \\ &= \inf_{A \in \mathcal{A} \mid A(x) > 0} I(A(x), P(A)) = \inf_{A \in \mathcal{A}} I(A(x), P(A)). \quad \square \end{aligned}$$

Since ‘probable’ can be considered as a fuzzy modality in an adequate fuzzy logic in which $P(A)$ is regarded as the truth value of ‘ A is probable’ (see e.g. [14,13]), in such a logic $C(x)$ is the truth value of the proposition.

‘For all $A \in \mathcal{A}$, if x is in A then A is probable’.

2.2. Examples

A general name for points with maximal centrality serving as location estimators may be maximal centrality estimators. In general, there may be more than one maximally central point (e.g. several modes). All the maximally central points form the *central region*. Points with centrality at least α form the α -*central region*. The three basic examples of location estimators are particular instances of maximally central points. For clarity, we will present the univariate case first, then the general multivariate case. Let us define the following families of reference events.

For each $M > 0$, let $\mathcal{A}_M^{\text{ex}}$ be formed by all the events $A_{-1,t}^{\text{ex}}$ and $A_{1,t}^{\text{ex}}$ on \mathbb{R} given by

$$\begin{aligned} A_{-1,t}^{\text{ex}}(x) &= \frac{t-x}{2t} \cdot I_{[-t,t]}(x), \\ A_{1,t}^{\text{ex}}(x) &= \frac{t+x}{2t} \cdot I_{[-t,t]}(x), \end{aligned}$$

where $t \geq M$.

Let \mathcal{A}^{me} be formed by all the events

$$A_{-1,t}^{\text{me}} = I_{(-\infty, t]}, \quad A_{1,t}^{\text{me}} = I_{[t, \infty)},$$

where $t \in \mathbb{R}$.

Finally, let \mathcal{A}^{mo} be formed by all the indicator functions of singletons in \mathbb{R} .

Recall that the α -quantile of a random variable ξ is the set

$$Q_\alpha = \{x \in \mathbb{R} \mid P(\xi \leq x) \geq \alpha, P(\xi \geq x) \geq 1 - \alpha\}.$$

Points in $Q_{.5}$ are called medians of ξ .

The following result shows that the mode and median maximize specific fuzzy sets of central points.

Theorem 2.2. Let ξ be a random variable. Then, for suitable choices of a family of reference events \mathcal{A} and a level $\alpha \in (0, 1]$, the α -central region C_α is:

- (1) The set of all modes of ξ , if ξ is discrete (take $\mathcal{A} = \mathcal{A}^{\text{mo}}$, $\alpha = h(C)$).
- (2) The set of all medians of ξ (take $\mathcal{A} = \mathcal{A}^{\text{me}}$, $\alpha = h(C) = .5$).
- (3) The interquantile interval $[\min Q_\alpha, \max Q_{1-\alpha}]$ of ξ (take $\mathcal{A} = \mathcal{A}^{\text{me}}$, $\alpha \in (0, .5]$).

Proof. Let P be the distribution induced by ξ .

Proof of (1): Fix $x \in \mathbb{R}$. Then,

$$C(x) = \sup \{ \alpha \in (0, 1] \mid P(I_{\{y\}}) \geq \alpha \cdot I_{\{y\}}(x) \forall y \in \mathbb{R} \} = \sup \{ \alpha \in (0, 1] \mid P(\{x\}) \geq \alpha \} = P(\{x\}).$$

Clearly, for $\alpha = h(C)$, the α -central region is formed by all x maximizing $P(\{x\})$, i.e. the modes.

Proof of (2): It follows from (3), taking $\alpha = .5$.

Proof of (3): We have

$$\begin{aligned} C_\alpha &= \{x \in \mathbb{R} \mid P(I_{(-\infty, t]}) \geq \alpha \cdot I_{(-\infty, t]}(x), P(I_{[t, \infty)}) \geq \alpha \cdot I_{[t, \infty)}(x) \forall t \in \mathbb{R}\} = \{x \in \mathbb{R} \mid P(\xi \leq t) \geq \alpha \forall t \geq x, \text{ and} \\ &\quad P(\xi \geq t) \geq \alpha \forall t \leq x\} = \{x \in \mathbb{R} \mid P(\xi \leq x) \geq \alpha, P(\xi \geq x) \geq 1 - \alpha\}. \end{aligned}$$

Any such x is in the q -quantile of ξ for any $q \in [0, 1]$ such that $q \geq \alpha$ and $1 - \alpha \geq q$, that is, $q \in [\alpha, 1 - \alpha]$. Therefore,

$$C_\alpha \subset \bigcup_{q \in [\alpha, 1 - \alpha]} Q_q = [\min Q_\alpha, \max Q_{1-\alpha}].$$

The converse inclusion is similar. \square

The expectation of a bounded random variable is covered by this framework, too. If the random variable is unbounded but non-negative, a one-sided version of the same approach still yields the expectation as the smallest value having full membership in the fuzzy set of central points.

Theorem 2.3. *Let ξ be an integrable random variable. Then, for suitable choices of a family of reference events \mathcal{A} , the central region C_1 is:*

- (1) *The expectation of ξ , if $|\xi|$ is almost surely bounded by some constant M (take $\mathcal{A} = \mathcal{A}_M^{\text{ex}}$).*
- (2) *The interval $[E[\xi], \infty)$, if $\xi \geq 0$ (take $\mathcal{A} = \{A_{-1,t}^{\text{ex}}\}_{t>0}$).*

Proof. Proof of (1): Fix $t \geq M$ and let P denote the distribution of ξ . Let us solve for x the inequality $P(A_{-1,t}^{\text{ex}}) \geq A_{-1,t}^{\text{ex}}(x)$. Firstly,

$$P(A_{-1,t}^{\text{ex}}) = \int_{[-t,t]} \frac{t-x}{2t} P(dx) = \int_{\mathbb{R}} \frac{t-x}{2t} P(dx) = E\left[\frac{t-\xi}{2t}\right] = \frac{t-E[\xi]}{2t}.$$

Secondly, all values of x for which $A_{-1,t}^{\text{ex}}(x) = 0$ are trivial solutions. Finally, whenever $A_{-1,t}^{\text{ex}}(x) > 0$, we need to solve the inequality

$$\frac{t-E[\xi]}{2t} \geq \frac{t-x}{2t},$$

whose solutions are $x \geq E[\xi]$. Hence, the set of all solutions is

$$S_{-1,t} = (-\infty, -t] \cup [t, \infty) \cup [E[\xi], \infty).$$

Reasoning analogously, the set of all solutions of $P(A_{1,t}^{\text{ex}}) \geq A_{1,t}^{\text{ex}}(x)$ is

$$S_{1,t} = (-\infty, -t] \cup [t, \infty) \cup (-\infty, E[\xi]].$$

From the definition of C ,

$$C_1 = \bigcap_{t \geq M} (S_{-1,t} \cap S_{1,t}) = \bigcap_{t \geq M} ((-\infty, -t] \cup [t, \infty) \cup \{E[\xi]\}) = \{E[\xi]\}.$$

Proof of (2): Like before, we solve the inequality $P(A_{-1,t}^{\text{ex}}) \geq A_{-1,t}^{\text{ex}}(x)$, with trivial solutions $(-\infty, -t] \cup [t, \infty)$ if $A_{-1,t}^{\text{ex}}(x) = 0$. Elsewhere,

$$P(A_{-1,t}^{\text{ex}}) = \int_{[-t,t]} \frac{t-x}{2t} P(dx) = .5 \cdot P(\xi \in [-t, t]) - (2t)^{-1} E[\xi \cdot I_{\{\xi \in [-t, t]\}}]$$

and

$$A_{-1,t}^{\text{ex}}(x) = \frac{t-x}{2t},$$

therefore $P(A_{-1,t}^{\text{ex}}) \geq A_{-1,t}^{\text{ex}}(x)$ if and only if

$$t \cdot P(\xi \in [-t, t]) - E[\xi \cdot I_{\{\xi \in [-t, t]\}}] \geq t - x,$$

i.e., since $\xi \geq 0$,

$$x \geq E[\xi \cdot I_{\{\xi \leq t\}}] + t \cdot P(\xi > t) = E[\min(\xi, t)].$$

The right-hand side is increasing and, by the monotone convergence theorem, converges to $E[\xi]$ as $t \rightarrow \infty$. Therefore,

$$C_1 = \bigcap_{t>0} ((-\infty, -t] \cup [t, \infty) \cup [E[\min(\xi, t)], \infty)) = \bigcap_{t>0} [E[\min(\xi, t)], \infty) = [E[\xi], \infty). \quad \square$$

The identity $\alpha = h(C)$ in Theorems 2.2 and 2.3 means that the mode, the median and the expectation are the maximally central estimators for those specific families of reference events.

Let us present a generalization of these results to the multivariate setting. The unit sphere of \mathbb{R}^d is denoted by S^{d-1} , and the scalar product of two vectors u and v is denoted $u \cdot v$.

In the univariate case, as a consequence of its definition the median maximizes the quantity $\min\{P(\xi \leq x), P(\xi \geq x)\}$. That is, the least probability of a halfline with origin x is maximized when x is the median. Analogously, a multivariate generalization, the *halfspace median* [35], is defined to be the point or set of points maximizing the quantity

$$d_{HS}(x) = \inf_{u \in S^{d-1}} P(\{y \in \mathbb{R}^d | u \cdot y \geq u \cdot x\}),$$

the least probability of a halfspace with x in its boundary.

We will use the following families of reference events.

Fix $M > 0$. Let $\mathcal{A}_M^{\text{ex}}$ be formed by all the events $A_{u,t}^{\text{ex}}$ on \mathbb{R}^d given by

$$A_{u,t}^{\text{ex}}(x) = \frac{t + u \cdot x}{2t} \cdot I_{\{x | -t \leq u \cdot x \leq t\}}(x),$$

where $u \in S^{d-1}$ and $t \geq M$.

Let \mathcal{A}^{me} be formed by all the events

$$A_{u,t}^{\text{me}}(x) = \begin{cases} 1, & u \cdot x \geq t, \\ 0, & u \cdot x < t, \end{cases}$$

where $u \in S^{d-1}$ and $t \in \mathbb{R}$. Thus \mathcal{A}^{me} is formed by all (crisp) halfspaces in \mathbb{R}^d .

Note that the unit sphere S^0 in \mathbb{R} is the set $\{-1, 1\}$, justifying the notation in the univariate case.

For the mode, let \mathcal{A}^{mo} be the family of all indicator functions of singletons in \mathbb{R}^d .

Theorem 2.4. Let ξ be a random vector in \mathbb{R}^d . Then, for suitable choices of a family of reference events \mathcal{A} and a level $\alpha \in (0, 1]$, the α -central region C_α is:

- (1) The set of all modes of ξ , if ξ is discrete (take $\mathcal{A} = \mathcal{A}^{\text{mo}}$, $\alpha = h(C)$).
- (2) The expectation of ξ , if ξ is bounded with $|\xi| \leq M$ almost surely (take $\mathcal{A} = \mathcal{A}_M^{\text{ex}}$, $\alpha = h(C) = 1$).
- (3) The halfspace median of ξ (take $\mathcal{A} = \mathcal{A}^{\text{me}}$, $\alpha = h(C)$).

Proof. Proof of (1): Analogous to the one-dimensional proof in Theorem 2.2.(1).

Proof of (2): Fix $u \in S^{d-1}$. Analogously to the proof of Theorem 2.3.(1), but with $u \cdot t$ replacing t , we show that the solutions to the system of inequalities

$$P(A_{v,t}^{\text{ex}}) \geq A_{v,t}^{\text{ex}}(x), \quad t \geq M, \quad v = \pm u$$

are

$$\{x \in \mathbb{R}^d | u \cdot x = E[u \cdot \xi]\}.$$

Accordingly, taking into account that $E[u \cdot \xi] = u \cdot E[\xi]$,

$$C_1 = \bigcap_{u \in S^{d-1}} \{x \in \mathbb{R}^d | u \cdot x = u \cdot E[\xi]\} = \{x \in \mathbb{R}^d | \forall u \in S^{d-1}, u \cdot x = u \cdot E[\xi]\} = \{E[\xi]\}.$$

For the last identity, one inclusion is trivial; for the other inclusion, take $u = e_i$ for the members of the standard basis $\{e_i\}_{i=1}^d$.

Proof of (3): The statement can be rephrased as saying that a point maximizes C if and only if it maximizes d_{HS} . To prove it, we will just show that, in fact, $C = d_{HS}$. Set $H_{u,t} = \{y \in \mathbb{R}^d | u \cdot y \geq t\}$. With this notation,

$$d_{HS}(x) = \inf_{u \in S^{d-1}} P(H_{u,x}),$$

whereas

$$\begin{aligned} C(x) &= \sup\{\alpha \in (0, 1] | P(H_{u,t}) \geq \alpha \cdot I_{H_{u,t}}(x) \forall u \in S^{d-1}, t \in \mathbb{R}\} = \sup\{\alpha \in (0, 1] | u \in S^{d-1}, t \in \mathbb{R}, x \in H_{u,t} \Rightarrow P(H_{u,t}) \geq \alpha\} \\ &= \inf_{u \in S^{d-1}, t \in \mathbb{R}, x \in H_{u,t}} P(H_{u,t}). \end{aligned}$$

Since $x \in H_{u,u \cdot x}$, clearly $d_{HS}(x) \geq C(x)$. To prove the converse, it suffices to check that, whenever $x \in H_{u,t}$ for $u \in S^{d-1}$ and $t \in \mathbb{R}$, we have $P(H_{u,t}) \geq P(H_{u,u \cdot x})$. But, if $x \in H_{u,t}$, then $u \cdot x \geq t$, yielding

$$H_{u,u \cdot x} = \{y | u \cdot y \geq u \cdot x\} \subset \{y | u \cdot y \geq t\} = H_{u,t},$$

whence the inequality follows. \square

2.3. Some remarks on reference events

An important question is how to choose the family \mathcal{A} of reference events (see Fig. 1). For space reasons, in this paper we content ourselves with showing that some choices of \mathcal{A} lead to fuzzy sets of central points whose maxima are well-known location estimators. Results on which properties of \mathcal{A} translate into nice properties of C will be presented in forthcoming work. Further questions of interest include understanding the influence of \mathcal{A} in the statistical behavior of C and its maximal centrality estimators, and on the feasibility of computationally reasonable algorithms to calculate C .

As an example of the properties which may need to be imposed on \mathcal{A} , consider the situation when reference events are chosen to be crisp. Then C adopts the simpler form

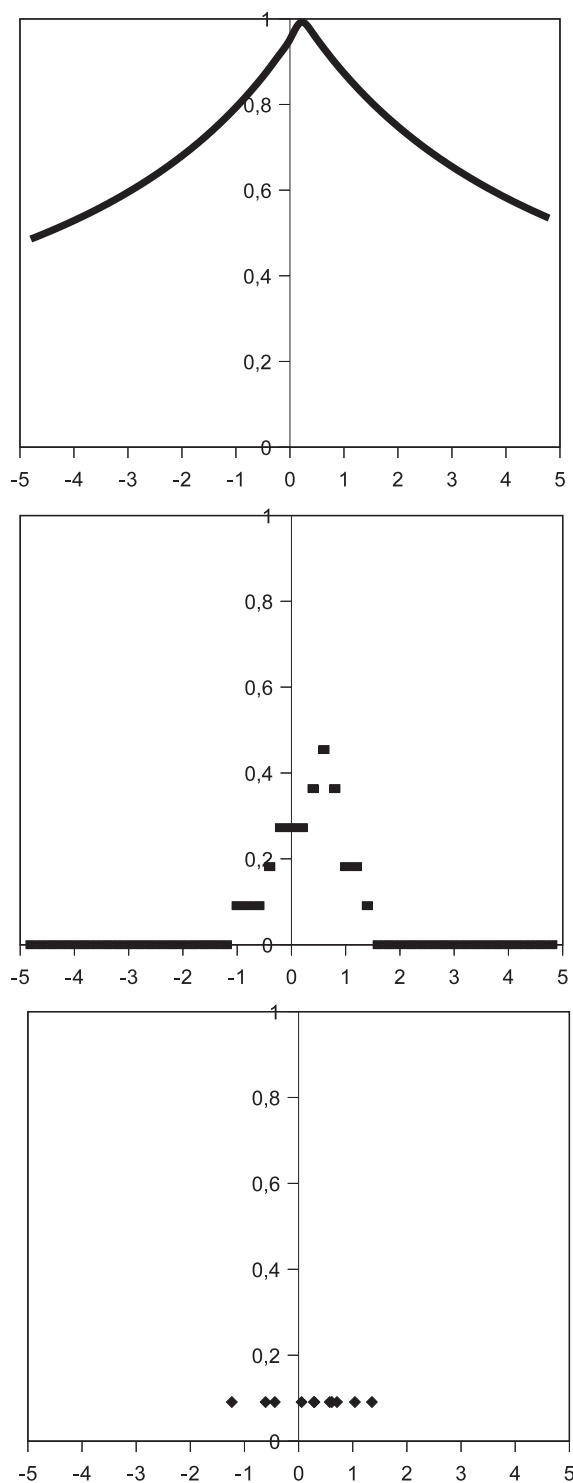


Fig. 1. Fuzzy sets of central points for an artificial sample from a standard normal distribution. The choice of reference events leads to having the sample mean, median and mode, respectively, as a maximal centrality estimator.

$$C(x) = \inf_{x \in A \in \mathcal{A}} P(A)$$

called by Zuo and Serfling [40] a *type D depth function*, and closely related to Small's *index functions* [29]. The conditions imposed by Zuo and Serfling are as follows:

- (i) \mathcal{A} is formed by closed (crisp) sets.
- (ii) The closure of the complement of each set in \mathcal{A} is also in \mathcal{A} .
- (iii) If x is an interior point of a set in \mathcal{A} , then there is a smaller set in \mathcal{A} for which x is a boundary point.

They showed that, under a technical condition, C is then upper semicontinuous with compact α -cuts, and is a convex fuzzy set if \mathcal{A} is formed by convex sets [40, Theorem 2.11]. They also proved a limit theorem for large samples, under further assumptions on \mathcal{A} [41, Theorem B.2].

Different families of reference events may lead to the same fuzzy set of central points. As an example showing that it may happen even for crisp events, take \mathcal{A} to be the family of all compact sets: it can be checked that C is the same as for \mathcal{A}^{mo} .

3. Extension to families of distributions

So far, we have studied the degree of centrality of a point in a probability distribution. To understand the advantage of extending that framework by defining the degree of centrality of a point in a family of probability distributions, denote by F^- the pseudoinverse of the distribution function of P and consider the *Lorenz curve* and the *nonscaled Lorenz curve*, given by

$$L(x) = \frac{\int_0^x F^-(x) dx}{E[\xi]}, \quad NL(x) = \int_0^x F^-(x) dx$$

and used in Economics to quantify the inequality or disparity in wealth and income distributions. The value of the Lorenz curve at x represents the proportion of the total wealth owned by the 100x% poorest individuals.

Finding the Lorenz curve is neither a problem of finding a point nor a location problem, yet it is subsumed by our framework as an example of the following general scheme:

- (i) Consider a probability distribution P .
- (ii) Define a suitable family \mathcal{P} of distributions related to P .
- (iii) Obtain the central or α -central region for that family, not P itself.
- (iv) That region codifies information about P , but no longer location information.

The choices of \mathcal{A} and α above yield an object describing location (the expectation) if applied to the distribution of ξ , while they yield an object describing inequality when applied to a related family of distributions (see the end of Section 4 for details).

Another example is the family of distributions $\{Q | Q \leq \alpha^{-1}P\}$ used in [4] to define trimmed regions for a probability P . The central region of that family constitutes a trimmed region for P (i.e. observed data outside that region may be labeled as outliers and trimmed out).

A different justification for the extension is that it allows us to encompass models where more than one probability distribution is considered, e.g. imprecise probability models. The scheme would then be:

- (i) Consider a model.
- (ii) Define the family \mathcal{P} of all distributions compatible with the model.
- (iii) Obtain the central or α -central region for that family.
- (iv) That region codifies location information about the underlying variable.

Examples of such models include: neighborhoods of a given distribution, contamination models, some parametric or non-parametric models, upper and lower probabilities, credal sets, Choquet capacities and random sets. When the model is given explicitly as a family of distributions, the application is trivial. We will explicitly work out the last two cases, in which the family of distributions is implicit.

A nice feature is that centrality-based statistical methods (i.e. using the available data to define a fuzzy set of central points and proceeding on the basis of the latter) would be no more complex for such situations than for ordinary random variables. It may be more expensive computationally to obtain the fuzzy set but, from that point on, the procedures would be exactly the same. The prospect of centrality-based methods is not unrealistic: similar methods exist, in the context of statistical depth functions, for exploratory data analysis, outlier detection, classification and many other problems (see e.g. [22]). Again, this topic is relegated to future work.

When a whole family of distributions is taken into account, a point x should be considered central in the family if it can be a central point of the true distribution. That leads to one of the following two approaches:

- (i) Call x central in \mathcal{P} if it is central in at least one distribution in \mathcal{P} , i.e. x is potentially central in the underlying distribution.
- (ii) Call x central in \mathcal{P} if it is central in all distributions in \mathcal{P} , i.e. x is guaranteedly central in the underlying distribution.

(As the reader may perceive, there is some analogy with upper and lower probabilities.) In this paper, we explore only the first view. The second one, suggested by Didier Dubois, may be equally meaningful in applications.

Accordingly, we have the following definition: given a family \mathcal{P} of probability distributions, its fuzzy set C of central points (with respect to a family \mathcal{A} of reference events) is defined by

$$C(x) = \sup\{\alpha \in (0, 1] \mid \forall A \in \mathcal{A}, \sup_{P \in \mathcal{P}} P(A) \geq \alpha \cdot A(x)\}.$$

The rest of the definitions (α -central region, etc.) are modified analogously. In the second view similar to a lower probability, an infimum would replace the supremum.

Observe that C , as a summary of the available information, is still a fuzzy subset of \mathbb{R}^d and so its complexity was not increased by passing from a single distribution to a family of distributions. The scheme presented as a second justification applies to many types of models, by considering the family of all probability distributions compatible with the model.

4. Random sets

When a random set is considered, the statistical model takes the following form. The underlying random variable ξ is not directly observed, only a larger set of values X containing ξ is known. Therefore, ξ is such that $\xi \in X$ almost surely, namely ξ is a selection of X . In that case, the implicit family of distributions \mathcal{P} contains those of all selections of X .

Most location estimators for random sets are unified as special cases of central or α -central regions, at least when the random set is bounded. The most popular notion of expectation for random sets is the *Aumann expectation* [20,3],

$$E_A[X] = \{E[\xi] \mid \xi \text{ integrable, } \xi \in X \text{ a.s.}\}.$$

We will consider also the *Herer expectation* [15], which depends on the chosen metric ρ :

$$E_H[X] = \{x \in \mathbb{R}^d \mid \forall y \in \mathbb{R}^d, \rho(x, y) \leq E[\rho_H(X, \{y\})]\},$$

where ρ_H is the Hausdorff metric given by

$$\rho_H(A, B) = \max\{\sup_{x \in A} \inf_{y \in B} \rho(x, y), \sup_{y \in B} \inf_{x \in A} \rho(x, y)\}.$$

The Herer expectation can be used when the Euclidean distance is deemed inappropriate for the data at hand, as is the case with compositional data [1].

Theorem 4.1. *Let X be a random compact set in \mathbb{R}^d such that $\|X\| \leq M$ almost surely for some $M > 0$. Let \mathcal{P} be the family of all its selectable distributions. Then, for appropriate choices of \mathcal{A} , the central region C_1 with respect to \mathcal{P} is:*

- (1) *The Herer expectation of X with respect to the metric ρ induced by an arbitrary norm $\|\cdot\|$.*
- (2) *The Aumann expectation of X , if Ω is non-atomic or X is almost surely convex. In general, the convex hull of the Aumann expectation (take $\mathcal{A} = \mathcal{A}_M^{\text{ex}}$).*

Proof. Proof of (1): Since all norms in \mathbb{R}^d are equivalent, there is some constant $R > 0$ such that

$$R^{-1} \cdot \|x\| \leq \|\cdot\| \leq R \cdot \|x\|$$

for all $x \in \mathbb{R}^d$. Accordingly, $\|X\| \leq M R$ almost surely and so $\|\xi\| \leq M R$ a.s. for each selection ξ of X . Let \mathcal{A}_M be the family formed by the following reference events $A_{t,y}$:

$$A_{t,y}(x) = \begin{cases} \frac{\rho(x,y)}{(R+1)t}, & \rho(x,y) \leq (R+1)t, \|\cdot\| \leq t, \\ 0, & \text{otherwise,} \end{cases}$$

where $t \geq M$ and $y \in \mathbb{R}^d$. Let C be the fuzzy set of central points of \mathcal{P} with respect to \mathcal{A}_M .

A point $x \in \mathbb{R}^d$ is in C_1 when

$$\sup_{P \in \mathcal{P}} P(A_{t,y}) \geq A_{t,y}(x) \forall t \geq M, \quad y \in \mathbb{R}^d.$$

Fix $t \geq M$. Then, the inequality holds trivially as soon as $\|\cdot\| > t$. If $\|\cdot\| \leq t$ then, by the definition of \mathcal{P} , the left-hand side is

$$\sup_{P \in \mathcal{P}} P(A_{t,y}) = \sup_{\xi \in X \text{ a.s.}} \int_{\rho(\xi,y) \leq (R+1)t} \frac{\rho(\xi,y)}{(R+1)t} dP = \sup_{\xi \in X \text{ a.s.}} \int \frac{\rho(\xi,y)}{(R+1)t} dP$$

(for the last identity, observe that $\rho(\xi,y) \leq \|\xi\| + \|y\| \leq (R+1)t$ almost surely).

Taking that into account, for fixed $t \geq M$ the point x satisfies the inequalities for all $y \in \mathbb{R}^d$ if and only if

$$\sup_{\xi \in Xa.s.} \int \frac{\rho(\xi, y)}{(R+1)t} dP \geq \frac{\rho(x, y)}{(R+1)t}$$

for all y such that $\|y\| \leq t$ and $\rho(x, y) \leq (R+1)t$. Therefore, the solution set includes the set

$$\{x \in \mathbb{R}^d | \rho(x, y) \leq \sup_{\xi \in Xa.s.} E[\rho(\xi, y)] \forall y \in \mathbb{R}^d | \|y\| \leq t, \rho(x, y) \leq (R+1)t\}.$$

For any pair (x, y) there exists a sufficiently large t for which both $\|y\| \leq t$ and $\rho(x, y) \leq (R+1)t$ hold. Accordingly,

$$\begin{aligned} C_1 &= \bigcap_{t \geq M} \{x \in \mathbb{R}^d | \rho(x, y) \leq \sup_{\xi \in Xa.s.} E[\rho(\xi, y)] \forall y \in \mathbb{R}^d | \|y\| \leq t, \rho(x, y) \leq (R+1)t\} \\ &= \{x \in \mathbb{R}^d | \forall y \in \mathbb{R}^d, \rho(x, y) \leq \sup_{\xi \in Xa.s.} E[\rho(\xi, y)]\}. \end{aligned}$$

To show that the right-hand side is exactly the Herer expectation of X , we just need to prove the identity

$$E[\rho_H(X, \{y\})] = \sup_{\xi \in Xa.s.} E[\rho(\xi, y)].$$

Note that

$$\rho_H(X, \{y\}) = \sup_{x \in X} \rho(x, y) = \max_{x \in X} \rho(x, y).$$

The supremum is reached because the function $\rho(\cdot, y)$ is continuous and X takes on compact values. The inequality \geq is then clear. For the converse, define

$$Y = \{x \in X | \rho(x, y) = \rho_H(X, \{y\})\},$$

which is a random set since it is the intersection of X and the ρ -sphere of center y and radius $\rho_H(X, \{y\})$. By the measurable selection theorem, Y has a selection ξ , for which trivially $E[\rho(\xi, y)] = E[\rho_H(X, \{y\})]$. Since $\xi \in X$, (1) is proven. Therefore, $C_1 = E_H[X]$ for the choice $\mathcal{A} = \mathcal{A}_M$.

Proof of (2): Since \mathcal{P} is formed by all selectionable distributions of X and

$$C_1 = \{x \in \mathbb{R}^d | \sup_{P \in \mathcal{P}} P(A) \geq A(x) \forall A \in \mathcal{A}_M^{\text{ex}}\},$$

by virtue of Theorem 2.4.(2) we have $E[\xi] \in C_1$ for all selections $\xi \in X$. Therefore, $E_A[X] \subset C_1$.

Conversely, assume $x \in C_1$, namely

$$\sup_{P \in \mathcal{P}} \int_{x | u \cdot x \in [-t, t]} \frac{t + u \cdot x}{2t} P(dx) \geq \frac{t + u \cdot x}{2x}$$

for all $t \geq M$ and $u \in S^{d-1}$. Equivalently,

$$\sup_{\xi \in Xa.s.} (t \cdot P(u \cdot \xi \in [-t, t]) + E[u \cdot \xi]) \geq t + u \cdot x \forall t, u.$$

By the Cauchy–Schwarz inequality,

$$|u \cdot \xi| \leq |\xi| \leq \|X\| \leq M \leq t,$$

whence the probability above is 1. Therefore, C_1 is the set of those x satisfying

$$u \cdot x \leq \sup_{\xi \in Xa.s.} E[u \cdot \xi].$$

By a classical result in [16],

$$\sup_{\xi \in Xa.s.} E[u \cdot \xi] = \sup_{y \in E_A[X]} u \cdot y.$$

But

$$C_1 = \{x \in \mathbb{R}^d | u \cdot x \leq \sup_{y \in E_A[X]} u \cdot y \forall u \in S^{d-1}\} = \text{co} E_A[X],$$

where the last identity follows from the separation theorem.

So far, it has been proved that

$$E_A[X] \subset C_1 \subset \text{co} E_A[X].$$

The identity $C_1 = \text{co} E_A[X]$ will follow if we show that C_1 is convex.

We start by observing that the reference events $A_{t,y} \in \mathcal{A}_M^{\text{ex}}$ are quasiconvex. Indeed, the lower level sets, for $\alpha \in (0, 1]$, are

$$\{x | A_{t,y}(x) \leq \alpha\} = \left\{x \mid \frac{t + u \cdot x}{2t} \leq \alpha\right\} \cap \{x | u \cdot x \in [-t, t]\} = \{x | u \cdot x \in [-t, \min\{t, (2\alpha - 1)t\}]\},$$

which is the intersection of two halfspaces and so a convex set.

Now take $x_1, \dots, x_k \in C_1$ and $\lambda_1, \dots, \lambda_k \in [0, 1]$ with $\sum_{i=1}^k \lambda_i = 1$. By the definition of C_1 ,

$$\sup_{P \in \mathcal{P}} (A_{t,u}) \geq A_{t,u}(x_i) \forall t, u, i,$$

whence

$$\sup_{P \in \mathcal{P}} (A_{t,u}) \geq \max_i A_{t,u}(x_i) \geq A_{t,u}\left(\sum_{i=1}^k \lambda_i x_i\right)$$

by the quasiconvexity. Since that is valid for all t, u , indeed $\sum_{i=1}^k \lambda_i x_i \in C_1$ and the proof that $C_1 = \text{co}E_A[X]$ is complete.

There only remains to note that the Aumann expectation is convex whenever X is convex or the probability space is nonatomic. In those cases, $C_1 = E_A[X]$. \square

To provide further examples, let us note that another notion covered by our framework is the *coverage function* (or one-point coverage function) of a random set X , which generalizes the probability mass function of a random variable and is given by $p_X(x) = P(x \in X)$.

Proposition 4.2. *Let X be a random closed set in \mathbb{R}^d . Then, p_X is the fuzzy set of central points of the family of all selectable distributions of X with respect to $\mathcal{A} = \mathcal{A}^{\text{mo}}$.*

For the proof, see [34, Proposition 5]. Accordingly, the probabilistic interpretation of a fuzzy set as the coverage function of a random set is a specific choice of a fuzzy set of central points for the random set [34].

As a consequence of Proposition 4.2, the Vorob'ev expectation and median of a random set [36,31], defined in the context of forest fire modeling and also used for medical imaging, are by definition α -central regions.

Proposition 4.3. *Let X be a random closed set in \mathbb{R}^d and \mathcal{P} the family of all its selections. Then, for appropriate choices of \mathcal{A} and $\alpha \in (0, 1]$, the α -central region C_α is:*

- (1) *The Vorob'ev expectation of X , if it exists (take $\mathcal{A} = \mathcal{A}^{\text{mo}}$, $\alpha = \inf\{\beta \in (0, 1] | E[\text{Leb}(X)] \leq \text{Leb}(\{x | p_X(x) \geq \beta\})\}$, where Leb denotes Lebesgue measure).*
- (2) *The Vorob'ev median of X , if it exists (take $\mathcal{A} = \mathcal{A}^{\text{mo}}$, $\alpha = .5$).*

Proof. It follows from Proposition 4.2 and the definition of the α -central region, for those specific values of α . \square

To close this section, we will show that the zonoid and lift zonoid of a distribution are subsumed by this framework. A general reference on the statistical applications of zonoids and lift zonoids is [25], where they are used to define a notion of statistical depth, with applications to trimmed regions, nonparametric tests of multivariate location and scale, multivariate measures of dispersion and stochastic orders.

The *zonoid* of an integrable random vector ξ is the convex body

$$Z(\xi) = \{E[g \cdot \xi] | g(\omega) \in [0, 1] \forall \omega \in \Omega\}.$$

The *lift zonoid* of ξ is the zonoid of the $(d + 1)$ -dimensional random vector (1ξ) .

Proposition 4.4. *Let ξ be a random vector in \mathbb{R}^d such that $|\xi| \leq M$ almost surely for some $M > 0$. Then, taking $\mathcal{A} = \mathcal{A}_M^{\text{ex}}$ and the adequate choice for \mathcal{P} , the central region C_1 of \mathcal{P} with respect to \mathcal{A} is:*

- (1) *The zonoid of ξ .*
- (2) *The lift zonoid of ξ .*

Proof. Proof of (1): The zonoid of ξ equals the Aumann expectation of the random segment $[0, \xi]$, see e.g. [25]. By Theorem 4.1.(2), taking \mathcal{P} to be the family of all selectable distributions of $[0, \xi]$ (i.e. the distributions induced by $g \cdot \xi$ for all measurable functions $0 \leq g \leq 1$) yields

$$C_1 = E_A[[0, \xi]] = Z(\xi).$$

Proof of (2): Apply part (1) to the random vector (1ξ) in \mathbb{R}^{d+1} . \square

In the univariate case, the lift zonoid is a symmetric convex body in \mathbb{R}^2 , having the nonscaled Lorenz curve NL as its lower boundary, and the dual curve $1 - NL$ as its upper boundary [18]. In other words, the Lorenz curve is the lower boundary of the lift zonoid of $\xi/E[\xi]$. Moreover, the $(d + 1)$ -volume of the lift zonoid is an inequality measure serving as a multivariate generalization of the Gini inequality index [19].

5. Choquet capacities

Let ν be a *Choquet capacity* in the sense of [17], namely a function from the Borel σ -algebra of \mathbb{R}^d to $[0, 1]$ with the following properties:

- (a) $\nu(\emptyset) = 0$, $\nu(\Omega) = 1$, $\nu(A) \leq \nu(B)$ if $A \subset B$,
- (b) $\nu(C_n) \searrow \nu(C)$ if $C_n \searrow C$ and C_n, C are closed,
- (c) $\nu(A_n) \nearrow \nu(A)$ if $A_n \nearrow A$.

A more general definition, in which closed sets are replaced by compact sets, is possible (e.g. [24, Chapter 1]). A capacity is called *2-alternating* if

$$\nu(A \cup B) + \nu(A \cap B) \leq \nu(A) + \nu(B),$$

and *infinitely alternating* if

$$\nu\left(\bigcap_{i=1}^n A_i\right) \leq \sum_{\emptyset \neq I \subset \{1, \dots, n\}} (-1)^{\text{card}(I)+1} \nu\left(\bigcup_{i \in I} A_i\right)$$

whenever A_1, \dots, A_n are Borel sets.

The *dual capacity* to ν is given by $\bar{\nu}(A) = 1 - \nu(A^c)$. The dual to a 2-alternating or infinitely alternating capacity has the property called *2-monotony* or *infinite monotony*, respectively.

The *Choquet integral* of a random variable ξ with respect to ν is

$$\int_{(C)} \xi d\nu = \int_0^\infty \nu(\{\xi \geq t\}) dt + \int_{-\infty}^0 [\nu(\{\xi \geq t\}) - 1] dt,$$

which exists provided at least one of the two improper Riemann integrals is finite. In that case, ξ is called ν -integrable.

Our notion of centrality can be applied to capacities by taking \mathcal{P} to be the core of ν , namely the family of all probability distributions $P \leq \nu$ dominated by ν (equivalently, $P \geq \bar{\nu}$).

Observe that Zadeh's definition of the probability of a fuzzy event extends immediately to capacities by setting $\nu(A) = \int_{(C)} A d\nu$.

Proposition 5.1. *Let I be Goguen's fuzzy implication, and let ν be a 2-alternating Choquet capacity. Then, taking \mathcal{P} to be the core of ν results in*

$$C(x) = \inf_{A \in \mathcal{A}} I(A(x), \nu(A)).$$

Proof. Since fuzzy events are bounded Borel functions and \mathbb{R}^d is a Polish space, combining parts (iii) and (iv) of Lemma A.2 in [5] yields

$$\nu(A) = \int_{(C)} A d\nu = \sup_{P \in \text{core}(\nu)} \int A dP = \sup_{P \in \text{core}(\nu)} P(A),$$

whence

$$C(x) = \sup\{\alpha \in (0, 1] \mid \nu(A) \geq A(x) \forall A \in \mathcal{A}\}.$$

The remainder of the proof mirrors that of Proposition 2.1. \square

The Choquet integrals with respect to ν and $\bar{\nu}$ are sometimes called the upper and lower Choquet integrals. They can be retrieved in our framework.

Theorem 5.2. *Let Ω be a Polish space, and ν an infinitely alternating Choquet capacity on its Borel sets. Let ξ be a random variable with $|\xi| \leq M$ almost surely for some $M > 0$. Let \mathcal{P} be the core of the induced capacity $\nu_\xi: A \mapsto \nu(\xi^{-1}(A))$. Then, for $\mathcal{A} = \mathcal{A}_M^{\text{ex}}$, the central region C_1 of \mathcal{P} is*

$$C_1 = \left[\int_{(C)} \xi d\bar{\nu}, \int_{(C)} \xi d\nu \right].$$

Proof. From [30, Corollary 3.2.6], Ω admits a maybe different Polish topology generating the same Borel σ -algebra and with respect to which ξ is continuous. With that new topology, by [23, Lemma 4] there exists a random compact set $X: [0, 1] \rightarrow \Omega$ such that $\text{core}(\nu)$ is the family of all selectionable distributions of X , and

$$E_A[\xi(X)] = \left[\int_{(C)} \xi d\bar{\nu}, - \int_{(C)} -\xi d\bar{\nu} \right].$$

We claim that \mathcal{P} is the family of all selectionable distributions of $\xi(X)$. Indeed, a distribution Q in \mathbb{R} is in the core of v_ξ if and only if

$$Q(A) \leq v_\xi(A) = v(\xi^{-1}(A))$$

for every closed set A (the inequality extends to all Borel sets by the regularity of Q). But, since v is the capacity associated to X ,

$$v(\xi^{-1}(A)) = P(X \cap \xi^{-1}(A) \neq \emptyset) = P(\xi(X) \cap A \neq \emptyset).$$

From [2, Theorem 2.1], the condition

$$Q(A) \leq P(\xi(X) \cap A \neq \emptyset) \forall A \text{ closed}$$

is necessary and sufficient for Q to be the distribution of a selection of $\xi(X)$. Finally, since

$$-\int_{(C)} \xi d\bar{v} = \int_{(C)} \xi dv,$$

an application of Theorem 4.1.(2) to $\xi(X)$ yields

$$C_1 = E_A[\xi(X)] = \left[\int_{(C)} \xi d\bar{v}, \int_{(C)} \xi dv \right].$$

6. Discussion

1. We believe the construction and interpretation of the fuzzy set C to be very natural within Fuzzy Set Theory. It may be unnecessary to comment that its present definition is computationally impractical when compared to the much simpler formulas individually known for the statistics unified in this paper. At the very least, the paper proposes a new characterization of those statistics, showing that they can be regarded as particular cases of a general framework rooted on fuzzy set theoretical notions.

Most often, statistical and probabilistic notions have been used to try to 'explain' fuzzy sets, while this paper explores the possibility of using fuzzy sets to give a unified view of various concepts from statistics. The prototypical example of a maximally central estimator is the mode, which corresponds to taking the crisp points as reference events. Thus, the paper shows that some weird statements like 'The median, the Lorenz curve and the Choquet integral are generalized modes' make definite, unexpected sense.

2. Some readers may object that the definition of the expectation as a maximally central estimator is circular, since it involves calculating the probabilities $P(A)$ for each reference event A ; A is a measurable mapping which can be conceived as a random variable, and then $P(A)$ is precisely its expectation $E[A]$. This criticism deserves consideration.

One reply is as follows. Both the probability of a fuzzy event and the expectation of a random variable are concrete interpretations of the integral of a function against a probability measure; 'fuzzy event' is not a fancy name for 'random variable', rather they represent different semantics of the mathematical object 'measurable function'.

The fact that two interpretations are possible does not necessarily make it sensible to mix them at the semantic level. It is not easy to make sense of the definition of a central region C_1 if we 'read' C and A as random variables and $P(A)$ as the expectation of A . To start with, the central region would be formed by the maximizers of the random variable C , quite obscure since Probability Theory does not attribute a meaning to the $\omega \in \Omega$ at which the value of a random variable is maximized. That is in sharp contrast with the fuzzy logic interpretation, according to which $x \in C_1$ if it makes the truth value of 'Reference events containing x are probable' as high as possible.

A more technical reply notes that, from the interpretive point of view, the key fact about $P(A)$ is that it involves an extension of the probability measure on ordinary events to fuzzy events, not that it is calculated by computing an integral. But it can be checked easily that, under some reasonable assumptions, the extension to fuzzy events given by $P(A) = \int A dP$ is unique.

More precisely, let Φ be a functional on fuzzy events with the following properties:

- (a) $\Phi(\alpha \cdot I_B) = \alpha \cdot P(B)$ for each Borel set $B \subset \mathbb{R}^d$,
- (b) $\Phi(\max(A, A')) = \Phi(A) + \Phi(A')$ whenever A, A' are fuzzy events and $\min(A, A') = 0$,
- (c) $\Phi(A_n) \nearrow \Phi(\sup_n A_n)$ whenever A_n is an increasing sequence.

Then, $\Phi = P$: use (a) and (b) to prove it for simple functions, and (c) to approximate an arbitrary fuzzy event by an increasing sequence of simple functions.

3. Another issue with the definition of the fuzzy set of central points is that it presumes that probabilities and degrees of membership are commensurate (since it involves expressions like ' $P(A) \geq A(x)$ '). The fuzzy logic interpretation in Proposition 2.1 provides a setting where both $P(A)$ and $A(x)$ share a common meaning.

Even if one does not adhere to that interpretation, adhering to property (a) above links probabilities and degrees of membership, since it implies that $P(\alpha) = \alpha$ (any degree of membership α must be the same thing as the probability of the fuzzy event with membership constant to α).

4. The expectation and related objects have been shown to be maximally central estimators for bounded random variables only (and it is necessary to know in advance a bound M). It is important to note that this is enough to include their sample versions with full generality. For instance, the mean of a sample $\{x_1, \dots, x_n\}$ is just the expectation of the random variable taking on values x_i with probability n^{-1} . Since the sample minimum and maximum are known, a bound M exists and is known.
5. The Herer and Aumann expectations are linked by the fact that the convex hull of the Aumann expectation is just the Herer expectation with respect to the Euclidean metric [33, Theorem 3]. Therefore, as an alternative to the fuzzy events in $\mathcal{A}_M^{\text{ex}}$ used in several of the results, one may use the reference events \mathcal{A}_M defined in the proof of Theorem 4.1.(1).
6. Fuzzy sets of central points are connected to the notion of statistical depth [22,25]. For instance, the function d_{HS} , shown to be a particular fuzzy set of central points in the proof of Theorem 2.4.(3), is *halfspace depth*. Analogously, α -central regions are related to depth-trimmed regions. Other connections can be found in [34] or will be pursued in further work.
7. Zadeh (e.g. [38]) considered how to apply the notion of *usuality* to the values of a random variable ξ , as it appears in natural language sentences like ‘Robert usually arrives home at 6 p.m.’. In his approach, a level-2 fuzzy set \mathfrak{U} of usual values is defined, and each fuzzy set A of possible values has membership in \mathfrak{U} given by

$$\mathfrak{U}(A) = E[A(\xi)].$$

One observation in point is that \mathfrak{U} is monotone in the sense that $A \subset A'$ implies $\mathfrak{U}(A) \leq \mathfrak{U}(A')$. It is not easy to see how to use \mathfrak{U} to make informative statements about ξ , as the fuzzy sets having larger \mathfrak{U} -membership are just those making a weaker restriction on the possible values of ξ .

One can use a fuzzy set of central values to describe the usual values of ξ . For instance, for the choice of reference events $\mathcal{A} = \mathcal{A}^{\text{me}}$ we have the following property: $P(\xi \in C_\alpha) \geq 1 - 2\alpha$. In the example above, if the .20-cut is [5.5, 6.5] that means that Robert arrives sooner than 5:30 at most 20% of the time, and later than 6:30 at most 20% of the time.

8. Some families presented in this paper, like $\mathcal{A}_M^{\text{ex}}$, involve fuzzy events indexed by all $t \geq M$. To compute the maximal centrality estimator practically, it would be enough to consider $t = M$. Indeed, the role of taking all $t > M$ is just to rule out the values $(-\infty, t] \cup [t, \infty)$ as solutions, but we know in advance that the estimator lies in $[-M, M]$ and so those points would never be considered in practice. It is also straightforward to consider a more efficient interval $[a, b]$ with known bounds a, b if, for instance, both a and b have the same sign.

Acknowledgements

This paper is an extended, rewritten version of the contribution presented at the Lisbon 2009 IFSA-EUSFLAT Conference. I am grateful to the guest editors for their selecting it for this special issue, particularly to Didier Dubois for a nice list of questions and suggestions for further work.

Research has been partially carried out while the author was with the Universidad de Zaragoza, and partially funded by Spain's Ministerio de Educacion y Ciencia (MTM2005-02254 and TSI2005-02511), Ministerio de Ciencia e Innovación (MTM2008-01519, TIN2008-06796-C04-04) and Aragón's Gobierno (PM2007-034).

References

- [1] J. Aitchison, The Statistical Analysis of Compositional Data, second ed., Blackburn, Caldwell, 2003.
- [2] Z. Artstein, Distributions of random sets and random selections, Israel J. Math. 46 (1983) 313–324.
- [3] R.J. Aumann, Integrals of set-valued functions, J. Math. Anal. Appl. 12 (1965) 1–12.
- [4] I. Cascos, M. López-Díaz, Integral trimmed regions, J. Multivariable Anal. 96 (2005) 404–424.
- [5] A. Castaldo, F. Maccheroni, M. Marinacci, Random correspondences as bundles of random variables, Sankhya: Indian J. Stat. 66 (2004) 409–427.
- [6] G. Coletti, R. Scozzafava, Conditional probability, fuzzy sets, and possibility: a unifying view, Fuzzy Set Syst. 144 (2004) 227–249.
- [7] G. de Cooman, A behavioural model for vague probability assessments, Fuzzy Set Syst. 154 (2005) 305–385 (with discussion).
- [8] D. Dubois, L. Foulloy, G. Mauris, H. Prade, Probability–possibility transformations, triangular fuzzy sets, and probabilistic inequalities, Reliab. Comput. 10 (2004) 273–297.
- [9] D. Dubois, E.P. Klement, R. Mesiar (Eds.), Fuzzy Sets, Probability, and Statistics – Gaps and Bridges, Johannes Kepler Universität, Linz. <<http://www.flll.uni-linz.ac.at/div/research/linz2007/LINZ2007Abstracts.pdf>>.
- [10] D. Dubois, H.T. Nguyen, H. Prade, Possibility theory, probability and fuzzy sets: misunderstandings, bridges and gaps, in: D. Dubois, H. Prade (Eds.), Fundamentals of Fuzzy Sets, 343–438, Kluwer, Boston, 2000.
- [11] C.J. Geyer, G.D. Meeden, Fuzzy and randomized confidence intervals and p -values, Stat. Sci. 20 (2005) 358–387 (with discussion).
- [12] G. González-Rodríguez, A. Colubi, M.A. Gil, A fuzzy representation of random variables: an operational tool in exploratory analysis and hypothesis testing, Comput. Stat. Data Anal. 51 (2006) 163–176.
- [13] P. Hájek, Complexity of fuzzy probability logics II, Fuzzy Set Syst. 158 (2007) 2605–2611.
- [14] P. Hájek, L. Godo, F. Esteva, Fuzzy logic and probability, in: Proceedings of the 11th Annual Conference Artificial Intelligence, 1995, pp. 237–244.
- [15] W. Herer, Mathematical expectation and martingales of random sub-sets of a metric space, Prob. Math. Stat. 11 (1990) 291–304.
- [16] F. Hiai, H. Umegaki, Integrals, conditional expectations, and Martin–Gales of multivalued functions, J. Multivariable Anal. 7 (1977) 149–182.
- [17] P.J. Huber, V. Strassen, Minimax tests and the Neyman–Pearson lemma for capacities, Ann. Stat. 1 (1973) 251–263.
- [18] G. Koshevoy, K. Mosler, The Lorenz zonoid of a multivariate distribution, J. Am. Stat. Assoc. 91 (1996) 73–882.
- [19] G.A. Koshevoy, K. Mosler, Multivariate Gini indices, J. Multivariable Anal. 60 (1997) 252–276.
- [20] H. Kudo, Dependent experiments and sufficient statistics, Nat. Sci. Rep. Ochanomizu Univ. 4 (1954) 151–163.
- [21] M. Laviolette, J.W. Seaman, J.D. Barrett, W.H. Woodall, A probabilistic and statistical view of fuzzy methods, Technometrics 37 (1995) 249–292 (with discussion).

- [22] R. Liu, R. Serfling, D. Souvaine (Eds.), *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, American Mathematical Society, Providence, 2006.
- [23] F. Maccheroni, M. Marinacci, A strong law of large numbers for capacities, *Ann. Prob.* 33 (2005) 1171–1178.
- [24] I. Molchanov, *Theory of Random Sets*, Springer, London, 2005.
- [25] K. Mosler, *Multivariate Dispersion, Central Regions and Depth. The Lift Zonoid Approach*, *Lect. Notes Statist.*, vol. 165, Springer, Berlin, 2002.
- [26] A. Neumaier, Clouds, fuzzy sets and probability intervals, *Reliab. Comput.* 10 (2004) 249–272.
- [27] T.J. Ross, J.M. Booker, W.J. Parkinson (Eds.), *Fuzzy Logic and Probability Applications: Bridging the Gap*, SIAM, Philadelphia, 2002.
- [28] N. Singpurwalla, J.M. Booker, Membership functions and probability measures of fuzzy sets, *J. Am. Stat. Assoc.* 99 (2004) 867–889 (with discussion).
- [29] C.G. Small, Measures of centrality for multivariate and directional distributions, *Can. J. Stat.* 15 (1987) 31–39.
- [30] S.M. Srivastava, *A Course on Borel Sets*, Springer, New York, 1998.
- [31] D. Stoyan, H. Stoyan, *Fractals, Random Shapes and Point Fields*, Wiley, Chichester, 1994.
- [32] P. Terán, Review of [28], *Math. Reviews* MR2090919, 2005, 2005k:60012.
- [33] P. Terán, On the equivalence of Aumann and Herer expectations of random sets, *Test* 17 (2008) 505–514.
- [34] P. Terán, Connections between statistical depth functions and fuzzy sets, in: C. Borgelt et al. (Eds.), *Combining Soft Computing and Statistical Methods in Data Analysis*, Springer, Berlin, 2010, pp. 611–618.
- [35] J.W. Tukey, Mathematics and the picturing of data, in: *Proceedings of the International Congress of Mathematicians*, Vancouver, B.C., 1974, pp. 523–531, *Canad. Math. Congress*, Montreal, 1975.
- [36] O.Y. Vorob'ev, Mean-Measure Modelling, Nauka, Moscow, 1984 (in Russian).
- [37] L.A. Zadeh, Probability measures of fuzzy events, *J. Math. Anal. Appl.* 23 (1968) 421–427.
- [38] L.A. Zadeh, Outline of a theory of usuality based on fuzzy logic, in: G.J. Klir, B. Yuan (Eds.), *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems. Selected Papers by Lotfi A. Zadeh*, 694–712, World Scientific, Singapore, 1996.
- [39] L.A. Zadeh, Toward a generalized theory of uncertainty (GTU) – an outline, *Inform. Sci.* 172 (2005) 1–40.
- [40] Y. Zuo, R. Serfling, General notions of statistical depth function, *Ann. Stat.* 28 (2000) 461–482.
- [41] Y. Zuo, R. Serfling, Structural properties and convergence results for contours of sample statistical depth functions, *Ann. Stat.* 28 (2000) 483–499.